SIAM J. MATRIX ANAL. APPL.
Vol. 27, No. 4, pp. 952–967

CONVERGENCE ANALYSIS OF A PAGERANK UPDATING ALGORITHM BY LANGVILLE AND MEYER*

ILSE C. F. IPSEN† AND STEVE KIRKLAND‡

Abstract. The PageRank updating algorithm proposed by Langville and Meyer is a special case of an iterative aggregation/disaggregation (SIAD) method for computing stationary distributions of very large Markov chains. It is designed, in particular, to speed up the determination of PageRank, which is used by the search engine Google in the ranking of web pages.

In this paper the convergence, in exact arithmetic, of the SIAD method is analyzed. The SIAD method is expressed as the power method preconditioned by a partial LU factorization. This leads to a simple derivation of the asymptotic convergence rate of the SIAD method.

It is known that the power method applied to the Google matrix always converges, and we show that the asymptotic convergence rate of the SIAD method is at least as good as that of the power method. Furthermore, by exploiting the hyperlink structure of the web it can be shown that the asymptotic convergence rate of the SIAD method applied to the Google matrix can be made strictly faster than that of the power method.

Key words. Google, PageRank, Markov chain, power method, stochastic complement, aggregation/disaggregation

AMS subject classifications. 15A51, 65F15, 65F10, 15A18, 15A42

DOI. 10.1137/S0895479804439808

1. Introduction.

The PageRank of a web page is an important ingredient in how the search engine Google determines the ranking of web pages [4, 25]. Computing PageRank amounts to computing the stationary distribution of a stochastic matrix whose size is now in the billions [21, 22]. Langville and Meyer [12, 13, 15] propose an updating algorithm to compute stationary distributions of large Markov chains, with the aim of speeding up the computation of PageRank.

The hyperlink structure of the web can be represented as a directed graph, whose vertices are web pages and whose edges are links. From it one can construct a stochastic matrix[1] P as follows. If web page i has $d_i \geq 1$ outgoing links, then for each link from web page i to page j, entry (i, j) of the matrix P is $1/d_i$. If there is no link from page i to page j, then entry (i, j) of P is 0. If page i has no outgoing links at all (a dangling node, or sink), then all elements in row i of P are set to $1/n$, where n is the order of P [24]. Thus P is a stochastic matrix, and in order to construct a primitive stochastic matrix, one forms the convex combination

$$G = cP + (1 - c)\mathbf{1}v^T,$$

where $\mathbf{1}$ is the column vector of all ones, v is a positive probability vector, the superscript T denotes the transpose, and $0 < c < 1$ a scalar. This is the Google matrix.

*Received by the editors January 9, 2004; accepted for publication (in revised form) by I. S. Dhillon February 16, 2005; published electronically February 16, 2006.
http://www.siam.org/journals/simax/27-4/43980.html

†Center for Research in Scientific Computation, Department of Mathematics, North Carolina State University, P.O. Box 8205, Raleigh, NC 27695-8205 (ipsen@math.ncsu.edu, http://www4.ncsu.edu/~ipsen/). This research was supported in part by NSF grants DMS-0209931 and DMS-0209695.

‡Department of Mathematics and Statistics, University of Regina, Regina, SK S4S 0A2, Canada (kirkland@math.uregina.ca, http://www.math.uregina.ca/~kirkland/). This research was supported in part by NSERC grant OGP0138251.

[1] A stochastic matrix P is a real square matrix with elements $0 \leq p_{ij} \leq 1$, whose rows sum to one, $\sum_j p_{ij} = 1$ for all i.

Element i of the stationary probability π of G represents the probability that a random surfer visits web page i and it is a major ingredient in the PageRank of page i.

Thus Google faces the problem of having to compute the stationary probability π of a very large stochastic matrix. Direct methods [29, section 2] are too time consuming. However, the matrix G is constructed to be primitive, and the power method applied to G converges, in exact arithmetic, to a multiple of π for any nonnegative starting vector [29, section 3.1]. The asymptotic convergence rate of the power method is given by the modulus of the second largest eigenvalue, in magnitude, of G. The original PageRank algorithm relied on the power method [25], but convergence can be slow [12, section 6]. Langville and Meyer present an updating algorithm to compute stationary probabilities of large Markov chains, by accelerating the power method with an aggregation/disaggregation step similar to [29, section 6.3]. We refer to their method as SIAD for *special iterative aggregation/disaggregation method* [12, 13, 15]. Numerical experiments illustrate the potential of the SIAD method in the context of PageRank computation [12, 15]. Langville and Meyer's SIAD method is a special case of more general iterative aggregation/disaggregation methods that can accommodate general aggregation and disaggregation operators [17, 18]. Here, however, we focus on the version by Meyer and Langville, whose restriction operator is an inexact stochastic complementation.

Related work. Our results for the Google matrix are built on the algorithm of Langville and Meyer [14, 12, 13, 15]. However, there are other approaches for speeding up the power method on the Google matrix. The following algorithms appear to work well in practice, but the convergence proofs, if available, are not rigorous. Extrapolation methods [2, 7, 11] construct a new iterate from a judicious linear combination of previous iterates where the methods in [2] were inspired by the rational expressions for the PageRank vector interms of c from [28] and further developed in [3]. Adaptive methods [11, 9] exploit the fact that pages with lower page rank tend to converge faster and do not recompute these components. A two-stage algorithm [16] computes stationary probabilities associated with dangling nodes and with nondangling nodes and then combines the two. It is shown that there are starting vectors for which this algorithm converges at least as fast as the power method. A three-stage algorithm [10] which exploits block structure in the web graph to partition the matrix appears to be an iterative aggregation/disaggregation method like those in [29].

Overview. We analyze, in exact arithmetic, the asymptotic convergence of the SIAD method. In particular we show that the SIAD method amounts to a power method preconditioned by a partial LU factorization (sections 3, 4, and 5) and give a simple derivation of the asymptotic convergence rate (section 5). We consider the choice of different stochastic complements for better convergence (section 6) and computable upper bounds on the asymptotic convergence rate (section 7). It is well known that the power method applied to the Google matrix G always converges; in this paper we show that the asymptotic convergence rate of the SIAD method applied to G is at least as good as that of the power method. We consider conditions that assure convergence of the SIAD method in general (section 8). At last we prove a stronger result for convergence of the SIAD method applied to G (section 9). By exploiting the hyperlink structure of the web we conclude that the asymptotic convergence rate of the SIAD method applied to G can be made strictly better than that of the power method. We also show how one can save computations when applying the SIAD method to G.

Notation. All vectors and matrices are real. The identity matrix is I with columns e_j, $\mathbf{1} = \begin{pmatrix}1 & \dots & 1\end{pmatrix}^T$ is the column vector of all ones. For a matrix P, the inequality $P \geq 0$ means that each element of P is nonnegative. Similarly, $P > 0$ means that each element of P is positive.

A stochastic matrix P is a square matrix with $P \geq 0$ and $P\mathbf{1} = \mathbf{1}$. A probability vector v is a column vector with $v \geq 0$ and $v^T\mathbf{1} = 1$. The one-norm of a column vector v is $\|v^T\| \equiv |v|^T\mathbf{1}$. The eigenvalues $\lambda_i(P)$ of P are labeled in order of decreasing magnitude, $|\lambda_1(P)| \geq |\lambda_2(P)| \geq \cdots$. The directed graph associated with P is denoted by $\Delta(P)$.

2. The problem. Let P be a stochastic matrix. We want to compute the stationary distribution π of P, that is, a column vector π such that [1, Theorem 2.(1.1)] $\pi^T P = \pi^T$ or $\pi^T(I - P) = 0$ where $\pi \geq 0$ and $\pi^T\mathbf{1} = 1$. If P is irreducible, then the eigenvalue 1 is simple and it is the spectral radius, and the stationary distribution $\pi > 0$ is unique [1, Theorem 2.(1.3)], [30, Theorem 2.1]. If P is primitive, then the eigenvalue 1 is distinct and all other eigenvalues have smaller magnitude [1, Theorem 2.(1.7)].

The matrix P may be large, so the idea is to work with only a small part of the matrix. That is, partition

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix},$$

where P_{11} and P_{22} are square. Ideally the dimension of P_{11} is small compared to the dimension of P. The methods considered here approximate the stationary distribution π of P by operating on matrices derived from P_{11} whose dimension is, essentially, the same as that of P_{11}.

3. Exact aggregation/disaggregation. We present an aggregation/disaggregation method that is the basis for the SIAD method [12, 13, 15], an updating algorithm designed to compute stationary distributions of large Markov matrices, in the context of Google's PageRank. Our presentation differs from traditional ones [12, 19, 29] because we focus on a 2×2 partitioning and we start from a block LDU decomposition. The aggregation algorithm computes the components of π as stationary distributions of smaller matrices.

Partition the irreducible stochastic matrix

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}, \qquad \pi = \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix}$$

so that P_{11} and P_{22} are square. Since $I - P$ is an M-matrix [1, Theorem 8.(4.2)], as well as singular and irreducible, the nontrivial leading principal submatrix $I - P_{11}$ is nonsingular [1, Theorem 6.(4.16)]. Hence we can factor $I - P = LDU$, where [19, proof of Theorem 2.3]

$$\begin{aligned} L &\equiv \begin{pmatrix} I & 0 \\ -P_{21}(I - P_{11})^{-1} & I \end{pmatrix}, \quad D \equiv \begin{pmatrix} I - P_{11} & 0 \\ 0 & I - S \end{pmatrix}, \\ U &\equiv \begin{pmatrix} I & -(I - P_{11})^{-1}P_{12} \\ 0 & I \end{pmatrix}, \quad S \equiv P_{22} + P_{21}(I - P_{11})^{-1}P_{12}, \end{aligned} \tag{3.1}$$

so that $I - S$ is the Schur complement of $I - P_{11}$ in $I - P$ [5, section 1]. The matrix S is also known as the stochastic complement of P_{22} in P [19, Definition 2.1]. It is a special case of a Perron complement in the context of nonnegative matrices [20, Definition 2.1].

Since U is nonsingular we have $\pi^T(I-P)=0$ if and only if $\pi^T LD = 0$. Hence

$$\pi_2^T S = \pi_2^T, \qquad \pi_1^T = \pi_2^T P_{21}(I-P_{11})^{-1}, \tag{3.2}$$

which means that π_2 is a stationary distribution for the smaller matrix S. The expressions (3.2) represent a partial version of the coupling theorem for 2×2 block matrices [19, Corollary 4.1].

Since P is irreducible and stochastic, so is S [19, Theorem 2.3]. Hence S has a unique, positive stationary distribution σ,

$$\sigma^T S = \sigma^T, \qquad \sigma^T \mathbf{1} = 1, \qquad \sigma > 0.$$

Therefore we can determine π_2 from the stationary distribution σ of S and then set $\pi_2 = \rho\sigma$ where the (as yet unknown) factor ρ is responsible for the normalization $\pi^T \mathbf{1} = 1$.

The component π_1 and the factor ρ can also be expressed as components of the stationary distribution of a smaller matrix, the "aggregated" matrix

$$A \equiv \begin{pmatrix} P_{11} & P_{12}\mathbf{1} \\ \sigma^T P_{21} & \sigma^T P_{22}\mathbf{1} \end{pmatrix}.$$

This is because (3.2), $\pi_2 = \rho\sigma$, and $\sigma^T \mathbf{1} = 1$ imply

$$\begin{pmatrix} \pi_1^T & \rho \end{pmatrix} (I - A) = 0, \qquad \begin{pmatrix} \pi_1^T & \rho \end{pmatrix} \mathbf{1} = 1.$$

Thus $\begin{pmatrix} \pi_1^T & \rho \end{pmatrix}$ is a stationary distribution of A. Since A is stochastic and irreducible [19, Theorem 4.1], it has a unique, positive stationary distribution α,

$$\alpha^T A = \alpha^T, \qquad \alpha^T \mathbf{1} = 1, \qquad \alpha > 0.$$

Uniqueness implies $\alpha^T = \begin{pmatrix} \pi_1^T & \rho \end{pmatrix}$.

We have shown that a block LDU decomposition of $\pi^T(I-P)=0$ leads to the expression of π_1 and π_2 as stationary distributions of two smaller matrices.

ALGORITHM 1 (exact aggregation/disaggregation). *Determine the stationary distribution π of a stochastic irreducible matrix P.*

1. *Determine the stationary distribution σ of $S \equiv P_{22} + P_{21}(I-P_{11})^{-1}P_{12}$.*
2. *Determine the stationary distribution α of $A \equiv \begin{pmatrix} P_{11} & P_{12}\mathbf{1} \\ \sigma^T P_{21} & \sigma^T P_{22}\mathbf{1} \end{pmatrix}$.*
3. *Partition $\alpha = \binom{\pi_1}{\rho}$, and set $\pi \equiv \binom{\pi_1}{\rho\sigma}$.*

The first two steps of Algorithm 1 can be interpreted as an aggregation because they take place on "aggregated" matrices of smaller size, while the third step represents a disaggregation that produces a long vector π of original size.

Since exact aggregation/disaggregation is too time consuming for large matrices, it makes sense to consider approximate aggregation/disaggregation methods. To understand what happens in the approximate aggregation/disaggregation method in section 4 we express π in terms of the block LDU decomposition (3.1).

PROPOSITION 3.1. *Let P be stochastic and irreducible. Then*

$$\pi^T = \rho \begin{pmatrix} * & \sigma^T \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix} L^{-1},$$

where the scalar ρ ensures $\pi^T \mathbf{1} = 1$, and $*$ represents an arbitrary vector of appropriate length.

Proof. This follows from (3.2) and Algorithm 1. ∎

4. Approximate aggregation/disaggregation. The approximate aggregation/disaggregation algorithm does away with the time-consuming computations involving the stochastic complement S. Instead of the exact stationary distribution σ of the stochastic complement S, we pick any positive probability vector $\tilde{\sigma}$, i.e., $\tilde{\sigma} > 0$ and $\tilde{\sigma}^T \mathbf{1} = 1$. The approximate aggregation matrix is

$$\tilde{A} \equiv \begin{pmatrix} P_{11} & P_{12}\mathbf{1} \\ \tilde{\sigma}^T P_{21} & \tilde{\sigma}^T P_{22}\mathbf{1} \end{pmatrix},$$

and it differs from the exact matrix A only in the last row.

The matrix $\tilde{A}$ is stochastic because P is stochastic. If P is irreducible, then so is $\tilde{A}$ is for the following reason. If P_{11} is of order k, then the directed graph $\Delta(\tilde{A})$ of $\tilde{A}$ contains $k+1$ vertices. The fact that $\tilde{\sigma}$ and $\mathbf{1}$ are positive vectors implies the following: $\Delta(\tilde{A})$ contains an edge from vertex $k+1$ to i, $1 \leq i \leq k$, if and only if $\Delta(P)$ contains an edge from vertex j, $j \geq k+1$, to i. Also, $\Delta(\tilde{A})$ contains an edge from vertex i, $1 \leq i \leq k$, to $k+1$ if and only if $\Delta(P)$ contains an edge from vertex i to j, $j \geq k+1$. Finally, $\Delta(\tilde{A})$ contains an edge from i to j, $1 \leq i, j \leq k$, if and only if $\Delta(P)$ contains an edge from vertex i to j. Since $\Delta(P)$ is strongly connected, $\Delta(\tilde{A})$ must be strongly connected as well. Hence $\tilde{A}$ is irreducible. Thus it has a unique, positive stationary distribution $\tilde{\alpha}$,

$$\tilde{\alpha}^T \tilde{A} = \tilde{\alpha}^T, \qquad \tilde{\alpha}^T \mathbf{1} = 1, \qquad \tilde{\alpha} > 0.$$

All by itself this approximate aggregation makes no progress. However following it up by a power method iteration is known to have "a very salutary effect" [29, section 6.3.1]. Below is a single iteration of the iterative aggregation/disaggregation algorithm. It makes up for the approximation $\tilde{\sigma}$ by appending one iteration of the power method with P to produce an approximation $\tilde{\pi}$ to π.

ALGORITHM 2 (approximate aggregation/disaggregation). *Determine an approximation $\tilde{\pi}$ to the stationary distribution π of a stochastic irreducible matrix P, in one iteration.*

1. *Select a vector $\tilde{\sigma}$ with $\tilde{\sigma} > 0$ and $\tilde{\sigma}^T \mathbf{1} = 1$.*
2. *Determine the stationary distribution $\tilde{\alpha}$ of $\tilde{A} \equiv \begin{pmatrix} P_{11} & P_{12}\mathbf{1} \\ \tilde{\sigma}^T P_{21} & \tilde{\sigma}^T P_{22}\mathbf{1} \end{pmatrix}$.*
3. *Partition $\tilde{\alpha} = \binom{\omega_1}{\tilde{\rho}}$, and set $\omega \equiv \binom{\omega_1}{\tilde{\rho}\,\tilde{\sigma}}$.*
4. *Multiply $\tilde{\pi}^T \equiv \omega^T P$.*

If it so happens that $\tilde{\sigma} = \sigma$, then according to Proposition 3.1, $\omega^T = \tilde{\pi}^T = \omega^T P$, and ω is the desired stationary distribution.

The approximate aggregation/disaggregation Algorithm 2 amounts, in exact arithmetic, to one iteration of the power method, preconditioned by a partial LU factorization.

REMARK 4.1. *Let P be stochastic and irreducible and $\tilde{\sigma}$ a vector with $\tilde{\sigma} > 0$ and $\tilde{\sigma}^T \mathbf{1} = 1$. Then*

$$\tilde{\pi}^T = \tilde{\rho} \begin{pmatrix} * & \tilde{\sigma}^T \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix} L^{-1} P,$$

where $\tilde{\rho}$ ensures $\tilde{\pi}^T \mathbf{1} = 1$.

Algorithm 2 can be viewed as applying one step of the power method, with the preconditioned matrix

$$\tilde{P} \equiv \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix} L^{-1} P,$$

to the vector $\begin{pmatrix} * & \tilde{\sigma}^T \end{pmatrix}$ and then normalizing it.

In the next section we execute Algorithm 2 repeatedly.

5. Iterative aggregation/disaggregation (SIAD). The iterative aggregation/disaggregation (SIAD) method [12, 13, 15] is an updating algorithm designed to compute PageRank. The method consists of repeated application of Algorithm 2. We show that the SIAD method can be viewed as a preconditioned power method and give a simple derivation of the asymptotic convergence rate.

ALGORITHM 3 (iterative aggregation/disaggregation (SIAD)). *Determine an approximation $\pi^{(k)}$ to the stationary distribution π of a stochastic irreducible matrix P, in k iterations.*

1. *Select a vector $\pi^{(0)} = \begin{pmatrix}\pi_1^{(0)} & \pi_2^{(0)}\end{pmatrix}$ with $\pi_2^{(0)} > 0$.*
2. *Do $k = 1, 2, \ldots$*
   (a) *Normalize $\sigma^{(k)} \equiv \pi_2^{(k-1)}/[\pi_2^{(k-1)}]^T\mathbf{1}$.*
   (b) *Determine the stationary distribution $\alpha^{(k)}$ of $A^{(k)} \equiv \begin{pmatrix} P_{11} & P_{12}\mathbf{1} \\ [\sigma^{(k)}]^T P_{21} & [\sigma^{(k)}]^T P_{22}\mathbf{1}\end{pmatrix}$.*
   (c) *Partition $\alpha^{(k)} = \begin{pmatrix}\omega_1^{(k)} \\ \rho_k\end{pmatrix}$, and set $\omega^{(k)} \equiv \begin{pmatrix}\omega_1^{(k)} \\ \rho_k\sigma^{(k)}\end{pmatrix}$.*
   (d) *Multiply $[\pi^{(k)}]^T \equiv [\omega^{(k)}]^T P$.*

Algorithm 3 is mathematically equivalent, in exact arithmetic, to the power method preconditioned by a partial LU factorization.

PROPOSITION 5.1. *Let P be stochastic and irreducible and $\pi^{(0)} = \begin{pmatrix}\pi_1^{(0)} & \pi_2^{(0)}\end{pmatrix}$ a vector with $\pi_2^{(0)} > 0$. Then Algorithm 3 produces iterates*

$$[\pi^{(k)}]^T = \tilde{\rho}\,[\pi^{(0)}]^T \tilde{P}^k, \qquad \text{where} \quad \tilde{P} \equiv \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix} L^{-1}P,$$

and $\tilde{\rho}$ ensures $[\pi^{(k)}]^T\mathbf{1} = 1$.

Proof. This follows from Remark 4.1. ∎

To determine the asymptotic convergence rate of Algorithm 3, label the eigenvalues $\lambda_i(P)$ of a matrix P in order of decreasing magnitude, $|\lambda_1(P)| \geq |\lambda_2(P)| \geq \cdots$.

THEOREM 5.2 (asymptotic convergence rate). *If P is stochastic and irreducible, then the asymptotic convergence rate of Algorithm 3 is $|\lambda_2(S)|$.*

Proof. According to Proposition 5.1, Algorithm 3 represents a power method with the matrix

$$\tilde{P} = \begin{pmatrix} 0 & 0 \\ P_{21}(I - P_{11})^{-1} & S \end{pmatrix}.$$

Because $\tilde{P}$ is block triangular, its nonzero eigenvalues are those of S. Since P is stochastic and irreducible, so is S [19, Theorem 2.3]. Hence its spectral radius is $|\lambda_1(S)| = 1$. Therefore the power method applied to $\tilde{P}$ has asymptotic convergence rate $|\lambda_2(S)|/|\lambda_1(S)| = |\lambda_2(S)|$. ∎

As a consequence, Algorithm 3 converges like the power method applied to the stochastic complement S. The example below illustrates the effect of the asymptotic convergence rate.

EXAMPLE 5.1. *Consider the stochastic irreducible matrix $P = aI_n + (1-a)\mathbf{1}\pi^T$ of order n, where $0 < a < 1$ is a scalar and π is a vector with $\pi^T\mathbf{1} = 1$ and $\pi > 0$. Then π is the stationary distribution of P. Partitioning gives*

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} = \begin{pmatrix} aI_m + (1-a)\mathbf{1}\pi_1^T & (1-a)\mathbf{1}\pi_2^T \\ (1-a)\mathbf{1}\pi_1^T & aI_{n-m} + (1-a)\mathbf{1}\pi_2^T \end{pmatrix}.$$

The stochastic complement is

$$S = P_{22} + P_{21}(I - P_{11})^{-1}P_{12} = aI_{n-m} + (1-a)\mathbf{1}\sigma^T,$$

where $\sigma^T = \pi_2^T/\pi_2^T\mathbf{1}$ *is the stationary distribution of* S*. The exact aggregation matrix is*

$$A = \begin{pmatrix} P_{11} & P_{12}\mathbf{1} \\ \sigma^T P_{21} & \sigma^T P_{22}\mathbf{1} \end{pmatrix} = \begin{pmatrix} aI_m + (1-a)\mathbf{1}\pi_1^T & (1-a)(\pi_2^T\mathbf{1})\,\mathbf{1} \\ (1-a)\pi_1^T & a + (1-a)(\pi_2^T\mathbf{1})\,\mathbf{1} \end{pmatrix}$$
$$= aI_{m+1} + (1-a)\mathbf{1}\alpha^T,$$

where $\alpha^T = \begin{pmatrix} \pi_1^T & \pi_2^T\mathbf{1} \end{pmatrix}$ *is the stationary distribution of* A*. The aggregation matrix* A *happens to have the same structure as the original matrix* P*.*

The subdominant eigenvalue of S *is* $\lambda_2(S) = a$*. Hence the asymptotic convergence rate of Algorithm 3 is* a*. What does this mean for the iterates* $\pi^{(k)}$*? Any positive probability vector* $\tilde{\sigma}$ *produces* $\tilde{A} = A$*; thus* $\alpha^{(k)} = \alpha$ *for all* k*. This means after one iteration the leading components of* $\pi^{(k)}$ *have converged, i.e.,* $\pi_1^{(k)} = \pi_1$ *for all* k*. As for the trailing components of* $\pi^{(k)}$*,* $[\pi_2^{(k)}]^T$ *is a multiple of* $[\pi_2^{(0)}]^T S^k$*, where* $S^k = a^k I_{n-m} + (1-a^k)\mathbf{1}\sigma^T$*. Hence*

$$[\sigma^{(k)}]^T = a^{k-1}\,\frac{[\pi_2^{(0)}]^T}{\gamma} + (1-a^{k-1})\,\frac{\pi_2^T}{\tilde{\rho}}, \qquad \tilde{\rho} \equiv \pi_2^T\mathbf{1}, \quad \gamma \equiv [\pi_2^{(0)}]^T\mathbf{1}.$$

The iterates are then

$$[\pi^{(k)}]^T = \begin{pmatrix} \pi_1^T & a^k\,\frac{\tilde{\rho}}{\gamma}[\pi_2^{(0)}]^T + (1-a^k)\pi_2^T \end{pmatrix}.$$

The leading component π_1 *is exact, and the trailing component* $a^k\,\frac{\tilde{\rho}}{\gamma}[\pi_2^{(0)}]^T + (1-a^k)\pi_2^T$ *shows clearly that the contribution of the initial vector* $\pi_2^{(0)}$ *decreases as a power of* $a = \lambda_2(S)$*, while the contribution of the target vector* π_2 *increases at the same rate.*

6. Stochastic complements. In this section we consider how different choices of stochastic complements affect the asymptotic convergence rate of Algorithm 3.

Remember that the asymptotic convergence rate of the power method applied to a primitive stochastic matrix P is $|\lambda_2(P)|$. In Theorem 5.2 we showed that the asymptotic convergence rate of Algorithm 3 applied to

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}$$

is $|\lambda_2(S)|$, where S is the stochastic complement of P_{22}. Unfortunately we cannot take for granted that Algorithm 3 has a better asymptotic convergence rate than the power method, i.e., that $|\lambda_2(S)| < |\lambda_2(P)|$. In fact, the opposite is possible, as we illustrate in one of the examples below.

One remedy for accelerating Algorithm 3 in case of slow convergence is to be more flexible about the choice of stochastic complements and not to limit oneself to stochastic complements from *leading* principal submatrices. One can form a stochastic complement from any principal submatrix [19, Definition 2.1]. One way to think about this (to keep the notation simple) is to apply a permutation similarity transformation to P [19, Lemma 2.1], i.e.,

$$QPQ^T = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix},$$

where Q is a permutation matrix, and use the stochastic complement of the permuted matrix

$$S \equiv P_{22} + P_{21}(I - P_{11})^{-1} P_{12}.$$

Note that a permutation similarity merely reorders the components of the stationary distribution and preserves the eigenvalues. That is, we can choose different stochastic complements by thinking of applying Algorithm 3 to a matrix QPQ^T that is permutationally similar to P. The question is, can we find permutations Q such that the stochastic complement S of QPQ^T has $|\lambda_2(S)| \leq |\lambda_2(P)|$? That is, can Algorithm 3 converge faster than the power method on a suitably reordered matrix?

The examples below illustrate different convergence possibilities for the power method and Algorithm 3.

EXAMPLE 6.1. *This example illustrates that Algorithm* 3 *can converge in one iteration, while the power method can fail to converge to the stationary distribution.*

Let v be a positive probability vector, and

$$P = \begin{pmatrix} 0 & v^T \\ \mathbf{1} & 0 \end{pmatrix}$$

a stochastic matrix. The eigenvalues of P are 1, −1, *and* 0. *Hence $|\lambda_2(P)| = 1$, which means that the power method applied to P fails to converge to the stationary distribution (unless the initial vector is the stationary distribution). The stochastic complement that arises from partitioning off the first row and column of P is $S = \mathbf{1}v^T$. Its eigenvalues are* 1 *and* 0. *Thus $|\lambda_2(S)| = 0$, so that Algorithm* 3 *converges in a single iteration.*

EXAMPLE 6.2. *This example illustrates that the power method and Algorithm* 3 *can have the same asymptotic convergence rate, regardless of the choice of stochastic complement.*

Consider the matrix from Example 5.1, $P = aI + (1-a)\mathbf{1}\pi^T$, *where π is a positive probability vector and $0 < a < 1$. The matrix P has just two distinct eigenvalues, $\lambda_1(P) = 1$ and $\lambda_i(P) = a$, $i > 1$. Each stochastic complement of P has the form $S = aI + (1-a)\mathbf{1}\sigma^T$, where σ is a positive probability vector. The eigenvalues of S are also* 1 *and a. In particular, $\lambda_2(S) = \lambda_2(P)$, so that the power method and Algorithm* 3 *converge at the same rate.*

EXAMPLE 6.3. *This example illustrates that the asymptotic convergence rate of Algorithm* 3 *can be significantly worse than that of the power method, regardless of the choice of stochastic complement.*

The irreducible stochastic matrix

$$P = \begin{pmatrix} 5/6 & 0 & 1/6 \\ 3/4 & 1/6 & 1/12 \\ 2/3 & 1/3 & 0 \end{pmatrix}$$

has eigenvalues $\lambda_1(P) = 1$ and $\lambda_2(P) = \lambda_3(P) = 0$. It turns out that 0 *is a defective eigenvalue, corresponding to a Jordan block of order* 2. *Thus the power method converges in at most two iterations for any initial probability vector.*

Now let us consider all possible stochastic complements that can arise during Algorithm 3. *Denote by S_i the stochastic complement that arises from P by simultaneously permuting row and column i to the first position and then forming the corresponding 2×2 stochastic complement. The stochastic complements have the following subdominant eigenvalues:*

$$\lambda_2(S_1) = -1/6, \qquad \lambda_2(S_2) = -2/15, \qquad \lambda_2(S_3) = 5/36.$$

In all cases $|\lambda_2(S_i)| > |\lambda_2(P)| = 0$, $1 \le i \le 3$. *Thus for any choice of stochastic complement, the asymptotic convergence rate of Algorithm* 3 *is slower than that of the power method.*

7. An upper bound on the asymptotic convergence rate. We discuss an upper bound on the subdominant eigenvalue $|\lambda_2|$, which also furnishes upper bounds for the asymptotic convergence rates of the power method and Algorithm 3. The upper bound can be easier to compute than $|\lambda_2|$ and has some properties convenient for a convergence analysis.

Example 6.3 shows that for any stochastic complement S of an irreducible stochastic matrix P, $|\lambda_2(S)| > |\lambda_2(P)|$ is possible, which means that for such a matrix the asymptotic convergence rate of Algorithm 3 is always slower than that of the power method. As far as Algorithm 3 is concerned, this situation is undesirable. However, for the standard upper bound τ on $|\lambda(S)|$ described below, it will transpire that $\tau(S) > \tau(P)$ is impossible. Denote by $\|v^T\| \equiv |v|^T \mathbf{1}$ the one-norm of a column vector v, and by e_j the jth column of the identity matrix. With

$$\tau(P) \equiv \frac{1}{2} \max_{i,j} \|(e_i - e_j)^T P\|,$$

we have, for any stochastic matrix P [27, section 4],

$$|\lambda_2(P)| \le \tau(P). \tag{7.1}$$

Observe that $\tau(P) \le 1$ for any stochastic matrix P. The quantity $\tau(P)$ represents an upper bound on the asymptotic convergence rate of the power method, while $\tau(S)$ is an upper bound on the asymptotic convergence rate of Algorithm 3. In contrast to the eigenvalues, where $|\lambda_2(S)| > |\lambda_2(P)|$ is possible, this cannot happen for the upper bounds τ. The following result shows that $\tau(S) \le \tau(P)$, when P is irreducible. This was already shown in [23, Theorem 3.1], but our proof is more direct. We also note that there is a one-line proof of this fact, which is based on rather more detailed knowledge of the properties of τ described in [27].

Proposition 7.1. *If*

$$P = \begin{array}{c} \\ k \\ n-k \end{array} \!\!\begin{array}{c} \begin{array}{cc} k & n-k \end{array} \\ \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} \end{array}, \qquad 1 \le k \le n-2,$$

is stochastic and irreducible, and $S \equiv P_{22} + P_{21}(I - P_{11})^{-1} P_{12}$, *then*

$$\tau(S) \le \tau(P).$$

Proof. We claim that $\|u^T (I - P_{11})^{-1} P_{12}\| \le \|u\|$ for *all* vectors u. To see that the claim is true, observe that $P\mathbf{1} = \mathbf{1}$ implies $(I - P_{11})^{-1} P_{12} \mathbf{1} = \mathbf{1}$. With $v_j \equiv |u_j|$, $1 \le j \le k$, the triangle inequality implies

$$|u^T (I - P_{11})^{-1} P_{12} e_i| \le v^T (I - P_{11})^{-1} P_{12} e_i, \qquad 1 \le i \le n-k.$$

Hence

$$\|u^T (I - P_{11})^{-1} P_{12}\| \le \sum_{i=1}^{n-k} v^T (I - P_{11})^{-1} P_{12} e_i = v^T (I - P_{11})^{-1} P_{12} \mathbf{1} = v^T \mathbf{1} = \|u\|,$$

which proves the claim.

For any two indices $1 \le i, j \le n - k$, the triangle inequality and the above claim imply

$$\begin{aligned} \|(e_i - e_j)^T S\| &\le \|(e_i - e_j)^T P_{22}\| + \|(e_i - e_j)^T P_{21}(I - P_{11})^{-1} P_{12}\| \\ &\le \|(e_i - e_j)^T P_{22}\| + \|(e_i - e_j)^T P_{21}\| = \|(e_{k+i} - e_{k+j})^T P\|. \end{aligned}$$

Hence

$$\tau(S) \le \frac{1}{2} \max_{k+1 \le i,j \le n} \|(e_i - e_j)^T P\| \le \tau(P). \qquad \square$$

Proposition 7.1 shows that the upper bound on the asymptotic convergence rate of Algorithm 3 is never worse than that of the power method. Note that Proposition 7.1 does not hold for $k = n - 1$ because S is 1×1.

The result below implies that there are matrices (including the Google matrix in section 9) for which the power method always converges, and for which the asymptotic convergence rate of Algorithm 3 is at least as good as that of the power method.

COROLLARY 7.2. *Let $G \equiv cP + (1-c)\mathbf{1}v^T$, where P is a stochastic matrix, $0 < c < 1$, and v is a positive probability vector. If $|\lambda_2(P)| = 1$, then for any stochastic complement S of G*

$$|\lambda_2(S)| \le |\lambda_2(G)| = c < 1.$$

Proof. Since v is positive and $c < 1$, G is irreducible. Inequality (7.1) and Proposition 7.1 imply $|\lambda_2(S)| \le \tau(S) \le \tau(G)$. The results [6, Theorem 2], [14, Theorem 4.1], [28, Theorem 2.3], and (7.1) imply $c = |\lambda_2(G)| \le \tau(G)$. At last, $\tau(G) \le c$ follows from $\tau(G) = c\tau(P) \le c$. $\square$

Note that the fact that $|\lambda_2(G)| = c$ in Corollary 7.2 was already shown in [6, Theorem 2] and [14, Theorem 4.1]. The crucial inequality in Corollary 7.2 is $|\lambda_2(S)| \le |\lambda_2(G)|$. The matrices in Examples 5.1 and 6.2 are special cases of the matrices in Corollary 7.2 with $P = I$.

For certain matrices, the upper bound τ provides an easily computable upper bound on the asymptotic convergence rate of Algorithm 3. These include matrices where P_{21} or P_{22} has a column with all elements positive.

REMARK 7.1. *If P is irreducible and stochastic and $p_{im} > 0$ for some m and $k + 1 \le i \le n$, then*

$$|\lambda_2(S)| \le 1 - \min_{k+1 \le i \le n} p_{im} < 1.$$

To see why this is true, note that rows i and j, $k + 1 \le i, j \le n$, have positive entries in position m and

$$\begin{aligned} \|(e_i - e_j)^T P\| &= \sum_{l=1}^{n} |p_{il} - p_{jl}| \le 1 - p_{im} + 1 - p_{jm} + |p_{im} - p_{jm}| \\ &= 2(1 - \min\{p_{im}, p_{jm}\}). \end{aligned}$$

Proposition 7.1 implies $|\lambda_2(S)| \le 1 - \min_{k+1 \le i \le n} p_{im} < 1$.

8. Conditions for convergence. We derive conditions that guarantee convergence, in exact arithmetic, of Algorithm 3.

The power method with a stochastic irreducible matrix P converges if $|\lambda_2(P)| < 1$, or equivalently if P is primitive [1, Definition 2.(1.8)]. However, just because P is primitive, a stochastic complement S is not necessarily primitive [19, section 5].

REMARK 8.1. *Let*

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}$$

be stochastic and irreducible. If P_{22} *is primitive or* P_{22} *has at least one nonzero diagonal element, then Algorithm* 3 *converges.*

This follows from [19, *Theorems* 5.1 *and* 5.2].

REMARK 8.2. *Primitivity of a stochastic complement* S *can also be related to properties of the directed graph* $\Delta(S)$*, which is also known as the* compressed directed graph [8, *sections* 1 *and* 4]. *A useful property of the compressed directed graph is the following: For any pair of indices* i, j *with* $1 \leq i, j \leq n-k$*, if there is a path from vertex* $k+i$ *to vertex* $k+j$ *in* $\Delta(P)$*, then there is a path from vertex* i *to vertex* j *in* $\Delta(S)$.

THEOREM 8.1. *If* P *is irreducible and stochastic, then the following holds:* P *has at least one primitive stochastic complement of order at least two* $\Longleftrightarrow$ P *is not a cyclic permutation matrix.*

Proof. First, we claim that P is a cyclic permutation matrix if and only if each column of P has at most one positive entry. One direction of the claim is clear. To see the other direction, suppose that each column of P has at most one positive entry. Because P is irreducible, each column must have exactly one positive entry. Further, since the entries of P are at most 1, and the sum of all entries in P is equal to the order of the matrix, the single positive entry in each column of P must be a 1. Thus, P is an irreducible stochastic $(0,1)$ matrix; hence P is a cyclic permutation matrix.

With the claim established, we now proceed with the proof of the result.

$\Longleftarrow$: Suppose that P is not a cyclic permutation matrix. Then some column m of P has at least two positive entries, i.e., $p_{i_1,m}, \ldots, p_{i_q,m} > 0$ for $i_1, \ldots, i_q$ with $q \geq 2$. With a suitable permutation Q, partition

$$QPQ^T = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}$$

so that P_{22} is $q \times q$, and P_{21} or P_{22} has a positive column. Remark 7.1 implies that $|\lambda_2(S)| < 1$, so that S is primitive.

$\Longrightarrow$: Suppose that

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}$$

is a cyclic permutation matrix. Since $\Delta(P)$ has exactly one cycle involving every vertex, and since $\Delta(P_{11})$ is a proper subgraph of $\Delta(P)$, $\Delta(P_{11})$ contains no cycles. Hence P_{11} is nilpotent, so $(I - P_{11})^{-1}$ is a finite sum of powers of P_{11}. In particular, $(I - P_{11})^{-1}$ is a matrix with nonnegative integer entries. Hence the stochastic complement $S = P_{22} + P_{21}(I - P_{11})^{-1}P_{12}$ is also a matrix of nonnegative integers. Since S is stochastic, it is a $(0,1)$ matrix, and since S is irreducible, it is a cyclic permutation matrix. Therefore S is not primitive. ∎

COROLLARY 8.2. *If the irreducible stochastic matrix* P *is not a cyclic permutation matrix, then* P *has at least one stochastic complement* S *for which Algorithm* 3 *converges, i.e.,* $|\lambda_2(S)| < 1$.

9. The Google matrix. We show that, when Algorithm 3 is applied to the Google matrix G with an appropriate choice of stochastic complement, then the asymptotic convergence rate is better than that of the power method. We also show how one can save computations when applying Algorithm 3 to G.

In Corollary 7.2 we showed that if $|\lambda_2(P)| = 1$, then $|\lambda_2(S)| \leq |\lambda_2(G)| = c < 1$ for any stochastic complement S. This means that the power method applied to G converges, and the asymptotic convergence rate of Algorithm 3 is at least as good as that of the power method. Now we prove a stronger result by exploiting the structure of the Google matrix and by choosing a particular stochastic complement. In this case the asymptotic convergence rate of Algorithm 3 is strictly better than that of the power method. We make the following assumptions; these assumptions are not restrictive because they are almost certainly satisfied by the Google matrix.

Assumptions 1.

1. The stochastic matrix P contains $k \geq 2$ essential index classes $C_1, \ldots, C_k$. The class of inessential indices (if any) is called D.
   Recall that an index i is inessential if for some index j there is a chain of links from page i to page j, but there is no chain of links back from page j to page i (i.e., $(P^m)_{ij} > 0$ for some power P^m, but $(P^k)_{ji} = 0$ for all powers P^k); otherwise the index i is essential. An essential index class C is a set of essential indices, such that there is a chain of links between any two indices in C, but no outgoing link to an index outside of C [26, section 1.2].
2. Each essential index class C_j contains an index i_j whose corresponding diagonal entry in P is 0; i.e., at least one web page in each C_j does not link back to itself.

These assumptions imply that the rows of P that are equal to $\frac{1}{n}\mathbf{1}^T$ all correspond to inessential indices in D. Also, $|\lambda_2(P)| = 1$ because P has at least 2 essential index classes. Note that class D contains not only dangling nodes, i.e., nodes without out-links, but also nodes that link to dangling nodes.

When P satisfies Assumptions 1 there is a permutation matrix Q that orders the rows and colums of QPQ^T into the following form: $i_1, \ldots, i_k, C_1 \backslash \{i_1\}, \ldots, C_k \backslash \{i_k\}, D$. Partition

$$QGQ^T = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix},$$

where G_{11} is $k \times k$ and corresponds to the indices $i_1, \ldots, i_k$, and let $S_G \equiv G_{22} + G_{21}(I - G_{11})^{-1} G_{12}$ be the corresponding stochastic complement.

THEOREM 9.1. *Let $G = cP + (1-c)\mathbf{1}v^T$ where $0 < c < 1$, v is a positive probability vector, and the stochastic matrix P satisfies Assumptions* 1. *Then*

$$|\lambda_2(S_G)| < |\lambda_2(G)|.$$

Proof. We exploit the structure of P, in several stages, to capture the structure of the stochastic complement S_G. Partition

$$QPQ^T = \begin{pmatrix} 0 & P_{12} \\ P_{21} & P_{22} \end{pmatrix},$$

where the leading diagonal block of order k is zero because the indices $i_1, \ldots, i_k$ belong to different essential classes, and each index corresponds to a zero diagonal element of P. While QPQ^T is presented as a 2×2 block matrix, whose leading principal block corresponds to indices $i_1, \ldots, i_k$, we will have occasion to further partition the second block into two classes, $C_1 \setminus \{i_1\}, \ldots, C_k \setminus \{i_k\}$, and D.

Partitioning

$$v = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$$

conformally with P gives, for the partition of QGQ^T,

$$G_{11} = (1-c)\mathbf{1}v_1^T, \qquad G_{12} = cP_{12} + (1-c)\mathbf{1}\begin{pmatrix} v_2^T & v_3^T \end{pmatrix},$$
$$G_{21} = cP_{21} + (1-c)\mathbf{1}v_1^T, \quad G_{22} = cP_{22} + (1-c)\mathbf{1}\begin{pmatrix} v_2^T & v_3^T \end{pmatrix}.$$

We claim that $|\lambda_2(S_G)| < c$, from which the conclusion follows with Corollary 7.2.

To establish this claim, write $S_G = G_{22} + G_{21}(I - G_{11})^{-1}G_{12}$ as a convex combination $S_G = cT + (1-c)\mathbf{1}z^T$, where

$$T \equiv P_{22} + P_{21}(I - G_{11})^{-1}G_{12}, \qquad z^T \equiv \begin{pmatrix} v_2^T & v_3^T \end{pmatrix} + v_1^T(I - G_{11})^{-1}G_{12}.$$

Using the three facts that $(I - G_{11})^{-1} = I + \beta\mathbf{1}v_1^T$, where

$$\beta \equiv \frac{1-c}{1-(1-c)\delta}, \qquad \delta \equiv v_1^T\mathbf{1};$$

the stochastic complement $P_{22} + P_{21}P_{12}$ of P is again a stochastic matrix; and $P_{12}\mathbf{1} = \mathbf{1}$, we conclude that T is stochastic and z is a positive probability vector. Since $\lambda_2(S_G) = c\lambda_2(T)$ [14, Theorem 4.1] it suffices to prove $|\lambda_2(T)| < 1$. To this end, we will show that T has at least one positive row and is irreducible. This implies T is primitive [1, Corollary 2.(4.8)], hence $|\lambda_2(T)| < 1$ [1, Definition 2.(1.8)].

The principal submatrix of P corresponding to the essential index class C_j can be written as

$$M_j = \begin{pmatrix} 0 & x_j^T \\ y_j & B_j \end{pmatrix},$$

where the leading row and column correspond to index i_j, and the remaining rows and columns correspond to the indices of $C_j \setminus \{i_j\}$. Since C_j is an essential index class, y_j is not the zero vector. Hence M_j is irreducible and stochastic, and so is its stochastic complement $B_j + y_j x_j^T$.

With

$$B \equiv \begin{pmatrix} B_1 & & \\ & \ddots & \\ & & B_k \end{pmatrix}, \quad X \equiv \begin{pmatrix} x_1^T & & \\ & \ddots & \\ & & x_k^T \end{pmatrix}, \quad Y \equiv \begin{pmatrix} y_1 & & \\ & \ddots & \\ & & y_k \end{pmatrix},$$

it follows that

$$QPQ^T = \begin{pmatrix} 0 & P_{12} \\ P_{21} & P_{22} \end{pmatrix} = \left(\begin{array}{c|cc} 0 & X & 0 \\ \hline Y & B & 0 \\ R_1 & R_2 & R_3 \end{array}\right).$$

Hence

$$T = \underbrace{P_{22} + cP_{21}P_{12}} + c\beta P_{21}\mathbf{1}v_1^T P_{12} + \underbrace{(1-c)(1+\beta\delta)P_{21}\mathbf{1}\begin{pmatrix} v_2^T & v_3^T \end{pmatrix}}.$$

Consider the last summand in T. The row $\begin{pmatrix} v_2^T & v_3^T \end{pmatrix}$ is positive because v is. Moreover

$$P_{21}\mathbf{1} = \begin{pmatrix} y_1 \\ \vdots \\ y_k \\ R_1\mathbf{1} \end{pmatrix},$$

where each vector y_j has at least one positive element. This means, T has at least one positive row corresponding to each $C_j \setminus \{i_j\}$, $1 \le j \le k$.

We still need to prove that T is irreducible. T being irreducible is equivalent to the directed graph $\Delta(T)$ being strongly connected [1, Theorem 2.(2.7)]. We will show that for any vertex in $\Delta(T)$ there is a path to any other vertex.

1. Paths within $C_j \setminus \{i_j\}$, $1 \le j \le k$: Consider the first two summands in T,

$$P_{22} + cP_{21}P_{12} = \begin{pmatrix} B + cYX & 0 \\ R_2 + cR_1X & R_3 \end{pmatrix}.$$

The leading principal submatrix is

$$B + cYX = \begin{pmatrix} B_1 + cy_1x_1^T & & \\ & \ddots & \\ & & B_k + cy_kx_k^T \end{pmatrix}.$$

Every diagonal element $B_j + cy_jx_j^T$ is irreducible because its directed graph is the same as that of the irreducible matrix $B_j + y_jx_j^T$, i.e., $\Delta(B_j + cy_jx_j^T) = \Delta(B_j + y_jx_j^T)$. Therefore $\Delta(T)$ contains a path from every vertex in $C_j \setminus \{i_j\}$ to all other vertices in $C_j \setminus \{i_j\}$.

2. Paths leaving $C_j \setminus \{i_j\}$, $1 \le j \le k$: The above fact that T contains at least one positive row for each $C_j \setminus \{i_j\}$ implies that $\Delta(T)$ contains a path from every vertex in $C_j \setminus \{i_j\}$ to any other vertex in $\Delta(T)$. Together with item 1, this implies $\Delta(T)$ contains a path from every vertex in $C_j \setminus \{i_j\}$ to any other vertex in $\Delta(T)$, $1 \le j \le k$.

3. Paths leaving D: From the leading two summands of T, $P_{22} + cP_{21}P_{12}$ with $c > 0$, we see that $\Delta(T)$ contains $\Delta(P_{22} + cP_{21}P_{12}) = \Delta(S_P)$, where

$$S_P \equiv P_{22} + P_{21}P_{12} = \begin{pmatrix} B & 0 \\ R_2 & R_3 \end{pmatrix} + \begin{pmatrix} Y \\ R_1 \end{pmatrix} \begin{pmatrix} X & 0 \end{pmatrix}$$

is a stochastic complement of P. Since every inessential index $d \in D$ has a path to some essential index class C_j, $\Delta(P)$ contains a path from d to every vertex in C_j. Remark 8.1 implies that therefore $\Delta(S_P)$ also contains a path from d to every vertex of $C_j \setminus \{i_j\}$. Since $\Delta(T)$ contains $\Delta(S_P)$, every vertex in D must have a path to some vertex of $C_j \setminus \{i_j\}$. With item 2, we conclude that $\Delta(T)$ contains a path from every vertex in D to any other vertex in D. ∎

Theorem 9.1 shows that for the Google matrix, there is always a choice of stochastic complement so that the asymptotic convergence rate of Algorithm 3 is strictly smaller than that of the power method. We give a simple illustration below.

EXAMPLE 9.1. *This example illustrates that the asymptotic convergence rate of Algorithm* 3 *applied to the Google matrix G can be the square of that of the power method.*

Let $G = cP + (1-c)\mathbf{1}v^T$, where $0 < c < 1$, v is a positive probability vector, and

$$P = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}.$$

If P is of order $n = 2k$, then its directed graph $\Delta(P)$ is a union of k cycles, each of length 2.

Partitioning $v^T = \begin{pmatrix} v_1^T & v_2^T \end{pmatrix}$ conformally with P gives, for the stochastic complement from Theorem 9.1,

$$S_G = (1-c)\mathbf{1}v_2^T + \left(cI + (1-c)\mathbf{1}v_1^T\right)\left(I - (1-c)\mathbf{1}v_1^T\right)^{-1}\left(cI + (1-c)\mathbf{1}v_2^T\right).$$

Multiplying out yields

$$S_G = c^2 I + \frac{1-c^2}{1-(1-c)\delta}\mathbf{1}(cv_1^T + v_2^T), \qquad \text{where} \quad \delta \equiv v_1^T\mathbf{1}.$$

The eigenvalues of S_G are 1 and c^2.

Hence the asymptotic convergence rate of Algorithm 3 for this matrix G is c^2, while the convergence rate of the power method is only c.

Below we illustrate that when Algorithm 3 is applied to the Google matrix G, the computation of the stationary vectors of the aggregated matrices $A^{(k)}$ is cheap.

REMARK 9.1. *Algorithm 3 applied to the Google matrix G involves the computation of stationary vectors of aggregated matrices*

$$A^{(k)} = \begin{pmatrix} G_{11} & G_{12}\mathbf{1} \\ [\sigma^{(k)}]^T G_{21} & [\sigma^{(k)}]^T G_{22}\mathbf{1} \end{pmatrix}.$$

For the stochastic complement S_G from Theorem 9.1, $G_{11} = (1-c)\mathbf{1}v_1^T$. Hence

$$A^{(k)} = \begin{pmatrix} (1-c)\mathbf{1}v_1^T & \gamma\mathbf{1} \\ [z^{(k)}]^T & 1 - [z^{(k)}]^T\mathbf{1} \end{pmatrix},$$

$$[z^{(k)}]^T \equiv [\sigma^{(k)}]^T G_{21}, \quad \gamma \equiv 1 - (1-c)\delta, \quad \delta \equiv v_1^T\mathbf{1}.$$

The simple structure of $A^{(k)}$ yields an explicit expression for its stationary vector,

$$\alpha^{(k)} = \frac{1}{\epsilon_k + \gamma}\begin{pmatrix} \epsilon_k(1-c)v_1^T + \gamma[z^{(k)}]^T & \gamma \end{pmatrix}, \qquad \text{where} \quad \epsilon_k \equiv [z^{(k)}]^T\mathbf{1}.$$

Thus, when Algorithm 3 is applied to the Google matrix G, the computation of the stationary vectors $\alpha^{(k)}$ is cheap.

Acknowledgments. We thank Amy Langville and Carl Meyer for many helpful discussions.

REFERENCES

[1] A. BERMAN AND R. J. PLEMMONS, *Nonnegative Matrices in the Mathematical Sciences*, Classics Appl. Math., SIAM, Philadelphia, 1994.

[2] C. BREZINSKI AND M. REDIVO-ZAGLIA, *The PageRank vector: Properties, computation, approximation and acceleration*, SIMAX, submitted.

[3] C. BREZINSKI, M. REDIVO-ZAGLIA, AND S. SERRA-CAPIZZANO, *Extrapolation Methods for PageRank Computations*, Comptes Rendus Acad. Sci., Sér. I, 340 (2005), pp. 393–397.

[4] S. BRIN AND L. PAGE, *The anatomy of a large-scale hypertextual web search engine*, Comput. Networks and ISDN Systems, 30 (1998), pp. 107–117.

[5] R. W. Cottle, *Manifestations of the Schur complement*, Linear Algebra Appl., 8 (1974), pp. 189–211.

[6] T. H. Haveliwala and S. D. Kamvar, *The Second Eigenvalue of the Google Matrix*, Technical report 2003-20, Stanford University, Stanford, CA, 2003, http://dbpubs.stanford.edu/pub/2003-20.

[7] T. H. Haveliwala, S. D. Kamvar, D. Klein, C. D. Manning, and G. H. Golub, *Computing PageRank Using Power Extrapolation*, Technical report 2003-45, Stanford University, Stanford, CA, 2003, http://dbpubs.stanford.edu/pub/2003-45.

[8] C. R. Johnson and C. Xenophontos, *Irreducibility and primitivity of Perron complements: Application of the compressed directed graph*, in Graph Theory and Sparse Matrix Computation, IMA Vol. Math. Appl. 56, A. George, J. R. Gilbert, and J. W. H. Liu, eds., Springer, New York, 1993, pp. 101–106.

[9] S. D. Kamvar, T. H. Haveliwala, and G. H. Golub, *Adaptive methods for the computation of PageRank*, Linear Algebra Appl., 386 (2004), pp. 51–65.

[10] S. D. Kamvar, T. H. Haveliwala, C. D. Manning, and G. H. Golub, *Exploiting the Block Structure of the Web for Computing PageRank*, Technical report SCCM03-02, Stanford University, Stanford, CA, 2003, http://www-sccm.stanford.edu/nf-publications-tech.html.

[11] S. D. Kamvar, T. H. Haveliwala, C. D. Manning, and G. H. Golub, *Extrapolation methods for accelerating PageRank computations*, in Proceedings of the Twelfth International World Wide Web Conference (WWW2003), Toronto, 2003, ACM Press, New York, pp. 261–270.

[12] A. N. Langville and C. D. Meyer, *Updating PageRank Using the Group Inverse and Stochastic Complementation*, Technical report CRSC-TR02-32, North Carolina State University, Raleigh, NC, 2002, http://www.ncsu.edu/crsc/reports/reports02.html.

[13] A. N. Langville and C. D. Meyer, *Updating the Stationary Vector of an Irreducible Markov Chain*, Technical report CRSC-TR02-33, North Carolina State University, Raleigh, NC, 2002, http://www.ncsu.edu/crsc/reports/reports02.html.

[14] A. N. Langville and C. D. Meyer, *Fiddling with PageRank*, Technical report CRSC-TR03-34, North Carolina State University, Raleigh, NC, 2003, http://www.ncsu.edu/crsc/reports/reports03.html.

[15] A. N. Langville and C. D. Meyer, *Updating the Stationary Vector of an Irreducible Markov Chain with an Eye on Google's PageRank*, Technical report, North Carolina State University, Raleigh, NC, 2004.

[16] C. P.-C. Lee, G. H. Golub, and S. A. Zenios, *A Fast Two-Stage Algorithm for Computing PageRank*, Technical report SCCM03-15, Stanford University, Stanford, CA, 2003, http://www-sccm.stanford.edu/nf-publications-tech.html.

[17] I. Marek and P. Mayer, *Convergence theory of some classes of iterative aggregation/disaggregation methods for computing stationary probability vectors of stochastic matrices*, Linear Algebra Appl., 363 (2003), pp. 177–200.

[18] I. Marek and D. B. Szyld, *Local convergence of the (exact and inexact) iterative aggregation method for linear systems and Markov operators*, Numer. Math., 69 (1994), pp. 61–82.

[19] C. D. Meyer, *Stochastic complementation, uncoupling Markov chains, and the theory of nearly reducible systems*, SIAM Rev., 31 (1989), pp. 240–272.

[20] C. D. Meyer, *Uncoupling the Perron eigenvector problem*, Linear Algebra Appl., 114/115 (1989), pp. 69–94.

[21] C. B. Moler, *The world's largest matrix computation*, MATLAB News and Notes, October, 2002.

[22] C. B. Moler, *Numerical Computing with MATLAB*, SIAM, Philadelphia, 2004.

[23] M. Neumann and J. Xu, *On the stability of the computation of the stationary probabilities of Markov chains using Perron complements*, Numer. Linear Algebra Appl., 10 (2003), pp. 603–618.

[24] A. Y. Ng, A. X. Zheng, and M. I. Jordan, *Link analysis, eigenvectors and stability*, in Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI), B. Nabel, ed., Morgan Kaufmann/Elsevier, San Francisco, 2001, pp. 903–910.

[25] L. Page, S. Brin, R. Motwani, and T. Winograd, *The PageRank Citation Ranking: Bringing Order to the Web*, 1999, http://dbpubs.stanford.edu/pub/1999-66.

[26] E. Seneta, *Non-negative Matrices and Markov Chains*, Springer-Verlag, New York, 1981.

[27] E. Seneta, *Explicit forms for ergodicity coefficients and spectrum localization*, Linear Algebra Appl., 60 (1984), pp. 187–197.

[28] S. Serra-Capizzano, *Jordan canonical form of the Google matrix: A potential contribution to the PageRank Computation*, SIAM J. Matrix Anal. Appl., 27 (2005), pp. 305–312.

[29] W. J. Stewart, *Introduction to the Numerical Solution of Markov Chains*, Princeton University Press, Princeton, NJ, 1994.

[30] R. S. Varga, *Matrix Iterative Analysis*, Prentice Hall, Englewood Cliffs, NJ, 1962.